DOCUMENT RESUME

ED 345 784 JC 920 264

AUTHOR Isonio, Steven
TITLE Judgments of Placement Writing Samples at Golden West College: An Evaluation of Inter-Rater Reliability.
INSTITUTION Golden West Coll., Huntington Beach, Calif.
PUB DATE Nov 91
NOTE 12p.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Statistical Data (110)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Community Colleges; Comparative Analysis; Correlation; Evaluation Methods; *Holistic Evaluation; *Interrater Reliability; *Rating Scales; *Student Evaluation; *Student Placement; Two Year Colleges; Writing (Composition); *Writing Tests

ABSTRACT

At Golden College (California), student writing samples are holistically scored by pairs of judges on a six-point scale. Judges are allowed to use plus and minus figures, thus converting the integer scale to a decimal scale of evaluation. In 1991, 499 writing samples written as part of the placement testing process for students in the Coast Community College District's SOAR program were analyzed for reliable scoring on the part of the judges. Two procedures were used to assess the extent to which judgments of writing samples were consistent. The first method entailed calculating the difference between the two ratings of all writing samples. For 34.4% of the pairs, the ratings from independent judges were identical. Ratings differed by one-third of a point in 30.9% of the cases, by two-thirds of a point in 15.4% of the cases, and by exactly one point in 13.4% of the cases. Overall, ratings differed by one point or less in 94.1% of the cases, exceeding the minimal standard of 90% recommended by the California community colleges (CCC). The second procedure used was a Pearson correlation coefficient which assessed the relationship between paired ratings for the writing samples. The correlation between ratings was moderately strong, and positive (r=.76), exceeding the minimal standard of .75 recommended by the CCC. The data supported the hypothesis that the ratings from independent judges were made in a consistent manner. Two references and an appendix of related data are attached. (JMC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED345784

Judgments of Placement Writing Samples at Golden West College: An Evaluation of Inter-Rater Reliability

GOLDEN WEST COLLEGE
November 1991

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

S. Isonio

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Steven Isonio, PhD
Advisor, Matriculation Research/Assessment

JC 920264

BEST COPY AVAILABLE

Judgments of Placement Writing Samples at Golden West College: An Evaluation of Inter-Rater Reliability

Steven Isonio, PhD

Background. A number of specific assessment validation requirements have been put forth regarding multiple-choice, objective assessment instruments (California Community Colleges, 1990). Additionally, colleges that use writing samples in their placement rules have an obligation to demonstrate that these essay tests meet certain minimal standards for reliability.

At Golden West College, as is typically the case, writing samples are holistically scored by pairs of judges. A 6-point scale is used with higher scores indicating a more thoughtful response to the theme topic and general mastery of most of the grammar and usage conventions of standard English. This scale is applied independently to the writing sample by both judges. In this context, reliability concerns the extent to which evaluations provided by different judges of the same writing sample are in agreement.

Method. A total of 499 writing samples written as a part of the placement testing process for students participating in the Coast Community College District SOAR program were used for the analysis. In accordance with established Golden West College procedures, a "norming" period in which judges of writing samples discuss standards and expectations and evaluate a small number of

samples preceded the reading and evaluation of the entire set of writing samples. Each writing sample was read by two judges who independently provided a rating. In cases where the ratings differ by more than one point (on the 6-point scale), an evaluation by a third reader was made. The rating of the writing sample is then typically used in conjunction with the objective test score to make a placement recommendation for the student. Figure 1 depicts the frequency distribution for all ratings applied to the writing samples in the present analysis (all Figures and Tables appear in the Appendix). As can be seen, the modal rating is "3", followed by "2"; the next most frequent ratings are "3-", "3+", and "4-".

The pairs of ratings of the samples were compiled for analysis. Since it is the practice of the judges to use "+" and "-" as a part of some ratings (e.g., a "3+" and "4-" are sometimes used to make fine distinctions between "3" and "4"), a conversion to decimals was necessary. Table 1 shows the values used for translating original ratings into converted ratings (with decimals) for purposes of the analysis.

Results. Two procedures were used to assess the extent to which judgments of writing samples are consistent. The first method entailed calculating the difference between the two converted ratings of all writing samples. A total of 499 pairs were analyzed. As Figure 2 indicates, for 34.4% of the pairs, the ratings from independent judges were identical. Ratings

differed by one-third of a point in 30.9% of the cases, by two-thirds of a point in 15.4% of the cases, and by exactly one point in 13.4% of the cases. Thus, in 94.1% of the cases, ratings differed by one-point or less. This exceeds the recommended minimal standard of 90% (California Community Colleges, 1990).

A Pearson correlation coefficient was calculated to assess the relationship between paired ratings for the writing samples, as described in Assessment Validation Project Local Research Options, Design 18 (Matriculation Assessment Work Group, 1991). The correlation between ratings for the 499 writing samples was moderately strong, and positive [$r = .76$, $p < .001$]. This value exceeds the recommended minimal standard of .75 (California Community Colleges, 1990).

Discussion. Evidence based upon writing samples produced during the placement testing portion of the SOAR program indicates that the ratings from independent judges are indeed made in a consistent manner. Both in terms of the proportion of pairs of ratings within one point of each other and the correlation between ratings, the degree of consistency exceeds the minimal standards specified by the California Community Colleges Chancellor's Office.

References

California Community Colleges (1990). Standards, Policies and Procedures for the Evaluation of Assessment Instruments Used in California Community Colleges.

Matriculation Assessment Work Group (1991). Assessment Validation Project Local Research Options.

APPENDIX

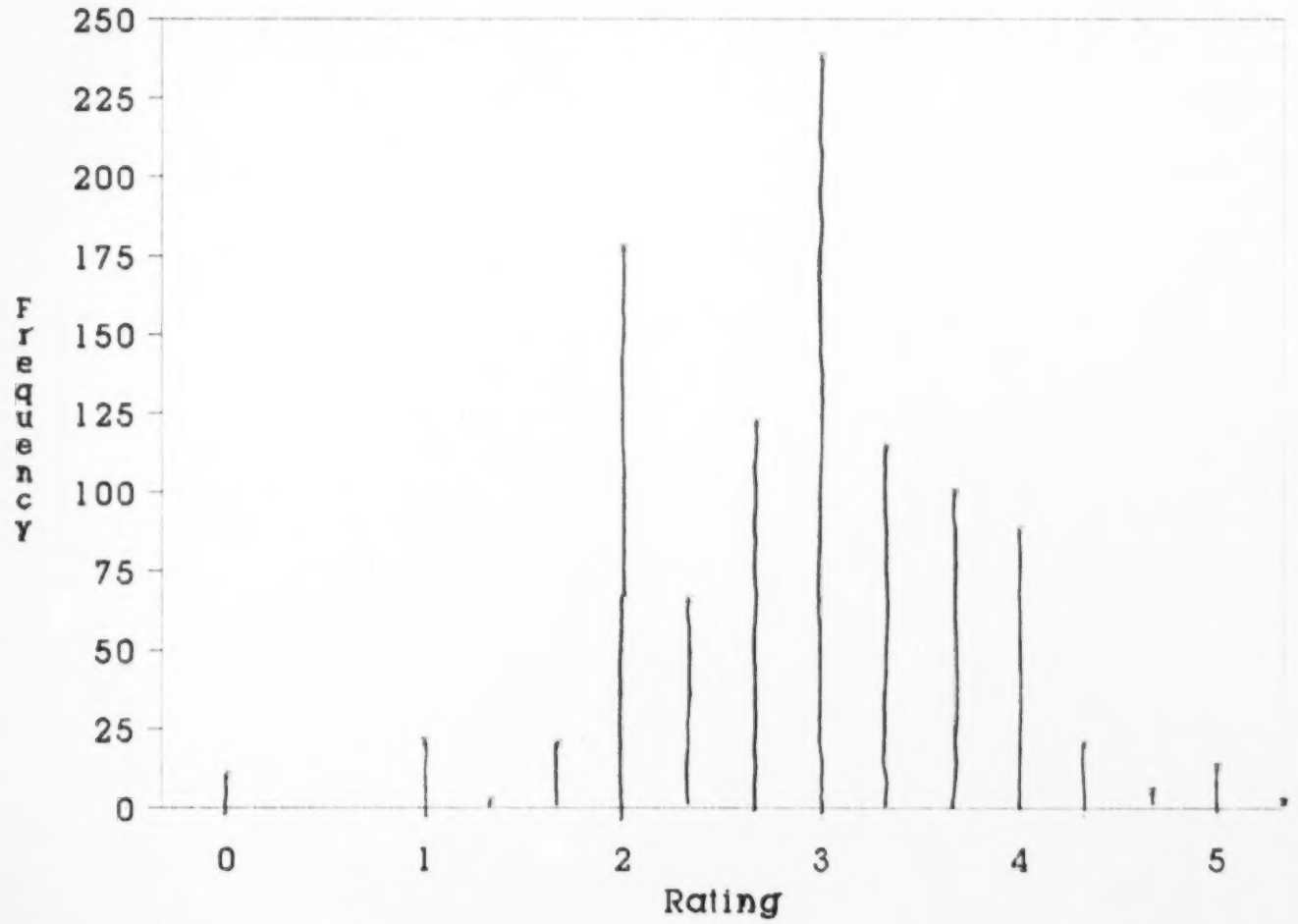
Table 1

Conversion From 6-point Scale to Decimal Scale

Rating Granted	Converted Rating
1-	0.67
1	1.00
1+	1.33
2-	1.67
2	2.00
2+	2.33
3-	2.67
3	3.00
3+	3.33
4-	3.67
4	4.00
4+	4.33
5-	4.67
5	5.00
5+	5.33
6-	5.67

Figure 1.

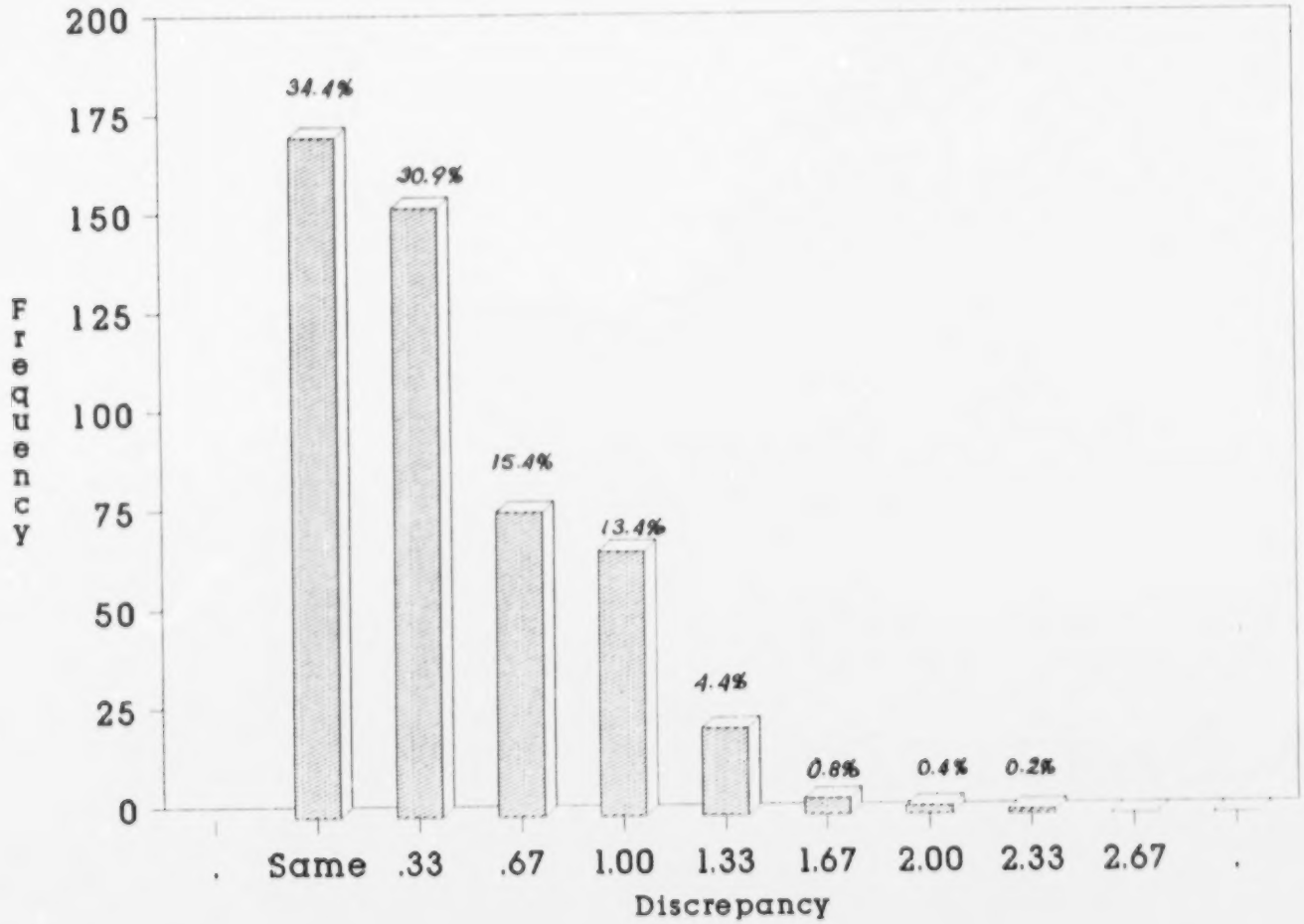
RATINGS OF PLACEMENT WRITING SAMPLES



1991 SOAR Students

Figure 2.

PWS Rating Discrepancies



SOAR 1991

ERIC Clearinghouse for
Junior Colleges

JUL 10 1992